CLAIMS

1. A similarity calculation device, which calculates an index for judging technical similarity between a first technical document group and a second technical document group, each comprising patent documents, technical reports, or other technical documents, characterized in comprising:

technical document group input means for inputting the first technical document group and the second technical document group for comparison;

technical information input means for inputting technical information such as keywords or IPC symbols;

cluster analysis means for retrieving technical documents containing the input technical information from technical documents contained in the first technical document group and the second technical document group, and for clustering the retrieved technical documents by each technical information;

similarity calculation means for calculating, as the similarity, the ratio of the number of intermixed clusters containing technical documents of both the first technical document group and the second technical document group, to the total number of clusters obtained as a result of the cluster analysis; and,

output means for outputting the calculated similarity to recording means, to display means, or to communication means.

2. A similarity calculation device, which calculates an index for judging technical similarity between a first

technical document group and a second technical document group, each comprising patent documents, technical reports, or other technical documents, characterized in comprising:

technical document group input means for inputting the first technical document group and the second technical document group for comparison;

technical information input means for inputting technical information such as keywords or IPC symbols;

cluster analysis means for retrieving technical documents containing the input technical information from technical documents contained in the first technical document group and the second technical document group, and for clustering the retrieved technical documents by each technical information;

similarity calculation means for calculating the total number of clusters obtained as a result of the cluster analysis and the number of intermixed clusters containing technical documents of both the first technical document group and the second technical document group, as well as for calculating the sum, over all intermixed clusters, of the product of a first correction value which takes a value according to the number of technical documents contained in each intermixed cluster and a second correction value which takes a value according to the state of mixing of technical documents of the first technical document group and the technical documents of the second technical document group in each intermixed cluster, and dividing the sum by the calculated total number of clusters to calculate the

similarity; and,

output means for outputting the calculated similarity to recording means, to display means, or to communication means.

3.   A similarity calculation device, which calculates an index for judging technical similarity between a first technical document group and a second technical document group, each comprising patent documents, technical reports, or other technical documents, characterized in comprising:

technical document group input means for inputting the first technical document group and the second technical document group for comparison;

technical information input means for inputting technical information such as keywords or IPC symbols;

cluster analysis means for retrieving technical documents containing the input technical information from technical documents contained in the first technical document group and the second technical document group, and for clustering the retrieved technical documents by each technical information;

similarity calculation means for calculating the total number of clusters obtained as a result of the cluster analysis and the number of intermixed clusters containing technical documents of both the first technical document group and the second technical document group, as well as for calculating the sum, over all intermixed clusters, of a correction value proportional to the αth power (where $0<\alpha$) of the number of technical documents in each cluster, and

dividing the sum by the calculated total number of clusters to calculate the similarity; and,

output means for outputting the calculated similarity to recording means, to display means, or to communication means.

4. A similarity calculation device, which calculates an index for judging technical similarity between a first technical document group and a second technical document group, each comprising patent documents, technical reports, or other technical documents, characterized in comprising:

technical document group input means for inputting the first technical document group and the second technical document group for comparison;

technical information input means for inputting technical information such as keywords or IPC symbols;

cluster analysis means for retrieving technical documents containing the input technical information from technical documents contained in the first technical document group and the second technical document group, and for clustering the retrieved technical documents by each technical information;

similarity calculation means for calculating the total number of clusters obtained as a result of the cluster analysis and the number of intermixed clusters containing technical documents of both the first technical document group and the second technical document group, as well as for calculating the sum, over all intermixed clusters, of a correction value obtained by dividing the αth power (where

0<α) of the number of technical documents in each cluster by a standardizing factor, and dividing the sum by the calculated total number of clusters to calculate the similarity; and,

output means for outputting the calculated similarity to recording means, to display means, or to communication means.

5. The similarity calculation device according to Claim 4, wherein the standardizing factor is the average value of the number of technical documents in all clusters.

6. A similarity calculation device, which calculates an index for judging technical similarity between a first technical document group and a second technical document group, each comprising patent documents, technical reports, or other technical documents, characterized in comprising:

technical document group input means for inputting the first technical document group and the second technical document group for comparison;

technical information input means for inputting technical information such as keywords or IPC symbols;

cluster analysis means for retrieving technical documents containing the input technical information from technical documents contained in the first technical document group and the second technical document group, and for clustering the retrieved technical documents by each technical information;

similarity calculation means for calculating the total number of clusters obtained as a result of the cluster

analysis and the number of intermixed clusters containing technical documents of both the first technical document group and the second technical document group, as well as for calculating the sum, over all intermixed clusters, of a correction value proportional to the γth power (where $0<\gamma$) of the probability of retrieving the m technical documents from the first technical document group and the n technical documents from the second technical document group, in order to perform correction according to the probability of the number of technical documents of the first technical document group and the second technical document group contained in each intermixed cluster obtained as a result of the cluster analysis, and dividing the sum by the calculated total number of clusters to calculate the similarity; and,

output means for outputting the calculated similarity to recording means, to display means, or to communication means.

7.   A similarity calculation device, which calculates an index for judging technical similarity between a first technical document group and a second technical document group, each comprising patent documents, technical reports, or other technical documents, characterized in comprising:

technical document group input means for inputting the first technical document group and the second technical document group for comparison;

technical information input means for inputting technical information such as keywords or IPC symbols;

cluster analysis means for retrieving technical documents containing the input technical information from technical documents contained in the first technical document group and the second technical document group, and for clustering the retrieved technical documents by each technical information;

similarity calculation means for calculating the total number of clusters obtained as a result of the cluster analysis and the number of intermixed clusters containing technical documents of both the first technical document group and the second technical document group, as well as for calculating the sum, over all intermixed clusters, of a correction value obtained by dividing, by a standardizing factor, the γth power (where $0<\gamma$) of the probability of retrieving the m technical documents from the first technical document group and the n technical documents from the second technical document group, in order to perform correction according to the probability of the number of technical documents of the first technical document group and the second technical document group contained in each intermixed cluster obtained as a result of the cluster analysis, and dividing the sum by the calculated total number of clusters to calculate the similarity; and,

output means for outputting the calculated similarity to recording means, to display means, or to communication means.

8.   The similarity calculation device according to Claim 7, wherein the standardizing factor is the γth power (where

0<γ) of the maximum value of the probability of retrieving the
m technical documents from the first technical document group
and the n technical documents from the second technical
document group.

9.   A similarity calculation device, which calculates an
index for judging technical similarity between a first
technical document group and a second technical document group,
each comprising patent documents, technical reports, or other
technical documents, characterized in comprising:

technical document group input means for inputting the
first technical document group and the second technical
document group for comparison;

technical information input means for inputting technical
information such as keywords or IPC symbols;

cluster analysis means for retrieving technical documents
containing the input technical information from technical
documents contained in the first technical document group and
the second technical document group, and for clustering the
retrieved technical documents by each technical information;

similarity calculation means for calculating the total
number of clusters obtained as a result of the cluster
analysis and the number of intermixed clusters containing
technical documents of both the first technical document group
and the second technical document group, as well as for
calculating the sum, over all intermixed clusters, of a
correction value proportional to the ζth power (where 0<ζ) of

130

the ratio of a composition ratio N/M and an intermixing ratio n/m, for the composition ratio N/M of the number of technical documents N contained in the second technical document group to the number of technical documents M contained in the first technical document group and for the intermixing ratio n/m of the number of technical documents n of the second technical document group to the number of technical documents m of the first technical document group contained in each intermixed cluster obtained as a result of the cluster analysis, and dividing the sum by the calculated total number of clusters to calculate the similarity; and,

output means for outputting the calculated similarity to recording means, to display means, or to communication means.

10. A similarity calculation device, which calculates an index for judging technical similarity between a first technical document group and a second technical document group, each comprising patent documents, technical reports, or other technical documents, characterized in comprising:

technical document group input means for inputting the first technical document group and the second technical document group for comparison;

technical information input means for inputting technical information such as keywords or IPC symbols;

cluster analysis means for retrieving technical documents containing the input technical information from technical documents contained in the first technical document group and

the second technical document group, and for clustering the retrieved technical documents by each technical information;

similarity calculation means for calculating the total number of clusters obtained as a result of the cluster analysis and the number of intermixed clusters containing technical documents of both the first technical document group and the second technical document group, and calculating an expectation value for retrieving a technical document of the first technical document group by multiplying the probability of retrieving a technical document of the first technical document group from among a technical document group covering the first technical document group and the second technical document group by the number of technical documents contained in each intermixed cluster, and calculating as an expectation value difference the difference between the expectation value and the number of technical documents of the first technical document group contained in each intermixed cluster, as well as for calculating the sum, over all intermixed clusters, of a correction value obtained by setting the expectation value difference as negative exponent for an arbitrary constant ξ (where $1<\xi$), and dividing the sum by the calculated total number of clusters to calculate the similarity; and

output means for outputting the calculated similarity to recording means, to display means, or to communication means.

11. A similarity calculation device, which calculates an index for judging technical similarity between a first

technical document group and a second technical document group, each comprising patent documents, technical reports, or other technical documents, characterized in comprising:

technical document group input means for inputting the first technical document group and the second technical document group for comparison;

technical information input means for inputting technical information such as keywords or IPC symbols;

cluster analysis means for retrieving technical documents containing the input technical information from technical documents contained in the first technical document group and the second technical document group, and for clustering the retrieved technical documents by each technical information;

similarity calculation means for calculating the total number of clusters obtained as a result of the cluster analysis and the number of intermixed clusters containing technical documents of both the first technical document group and the second technical document group, and calculating the expectation value for retrieving a technical document of the first technical document group by multiplying the probability of retrieving a technical document of the first technical document group from among a technical document group covering the first technical document group and the second technical document group by the number of technical documents contained in each intermixed cluster, and calculating as an expectation value difference the difference between the expectation value and the number of technical documents of the first technical

133

document group contained in each intermixed cluster, as well as for calculating the sum, over all intermixed clusters, of a correction value obtained by dividing the expectation value difference by the number of technical documents in each intermixed cluster and setting the divided expectation value difference as negative exponent for an arbitrary constant ξ (where $1<\xi$), and then dividing the sum by the calculated total number of clusters to calculate the similarity; and

output means for outputting the calculated similarity to recording means, to display means, or to communication means.

12. A similarity calculation program for calculating an index for judging technical similarity between technical document groups, which operates by means of information processing means for a similarity calculation device comprising technical document group input means for inputting the technical document groups, technical information input means for inputting technical information such as keywords, cluster analysis means for performing cluster analysis of the technical document groups by the technical information, similarity calculation means for calculating the total number of clusters and the number of intermixed clusters and calculating the similarity, output means for outputting the calculated similarity, and information processing means capable of controlling the technical document group input means, the technical information input means, the cluster analysis means, the similarity calculation means, and the

134

output means,

characterized in causing the information processing means to achieve:

a function, executed by the technical document group input means, for input of a first technical document group and a second technical document group for comparison;

a function, executed by the technical information input means, for input of the technical information such as keywords or IPC symbols;

a function, executed by the cluster analysis means, for retrieving technical documents containing the input technical information from technical documents contained in the first technical document group and the second technical document group, and for clustering the retrieved technical documents by each technical information;

a function, executed by the similarity calculation means, for calculating the total number of clusters obtained as a result of the cluster analysis and the number of intermixed clusters containing technical documents of both the first technical document group and the second technical document group, and for calculating, as the similarity, the ratio of the number of intermixed clusters, containing technical documents of both the first technical document group and the second technical document group, to the total number of clusters obtained as a result of the cluster analysis; and

a function, executed by the output means, for outputting the calculated similarity to recording means, to display means,

135

or to communication means.

    13.   A similarity calculation program for calculating an
index for judging technical similarity between technical
document groups, which operates by means of information
processing means for a similarity calculation device
comprising technical document group input means for inputting
the technical document groups, technical information input
means for inputting technical information such as keywords,
cluster analysis means for performing cluster analysis of the
technical document groups by the technical information,
similarity calculation means for calculating the total number
of clusters and the number of intermixed clusters and
calculating the similarity, output means for outputting the
calculated similarity, and information processing means
capable of controlling the technical document group input
means, the technical information input means, the cluster
analysis means, the similarity calculation means, and the
output means,

    characterized in causing the information processing means
to achieve:

    a function, executed by the technical document group
input means, for input of a first technical document group and
a second technical document group for comparison;

    a function, executed by the technical information input
means, for input of the technical information such as keywords
or IPC symbols;

a function, executed by the cluster analysis means, for retrieving technical documents containing the input technical information from technical documents contained in the first technical document group and the second technical document group, and for clustering the retrieved technical documents by each technical information;

a function, executed by the similarity calculation means, for calculating the total number of clusters obtained as a result of the cluster analysis and the number of intermixed clusters containing technical documents of both the first technical document group and the second technical document group, as well as for calculating the sum, over all intermixed clusters, of the product of a first correction value which takes a value according to the number of technical documents contained in each intermixed cluster and a second correction value which takes a value according to the state of mixing of technical documents of the first technical document group and the technical documents of the second technical document group in each intermixed cluster, and dividing the sum by the calculated total number of clusters to calculate the similarity; and

a function, executed by the output means, for outputting the calculated similarity to recording means, to display means, or to communication means.

14. A similarity calculation program for calculating an index for judging technical similarity between technical

document groups, which operates by means of information

processing means for a similarity calculation device

comprising technical document group input means for inputting

the technical document groups, technical information input

means for inputting technical information such as keywords,

cluster analysis means for performing cluster analysis of the

technical document groups by the technical information,

similarity calculation means for calculating the total number

of clusters and the number of intermixed clusters and

calculating the similarity, output means for outputting the

calculated similarity, and information processing means

capable of controlling the technical document group input

means, the technical information input means, the cluster

analysis means, the similarity calculation means, and the

output means,

characterized in causing the information processing means

to achieve:

a function, executed by the technical document group

input means, for input of a first technical document group and

a second technical document group for comparison;

a function, executed by the technical information input

means, for input of the technical information such as keywords

or IPC symbols;

a function, executed by the cluster analysis means, for

retrieving technical documents containing the input technical

information from technical documents contained in the first

technical document group and the second technical document

group, and for clustering the retrieved technical documents by each technical information;

a function, executed by the similarity calculation means, for calculating the total number of clusters obtained as a result of the cluster analysis and the number of intermixed clusters containing technical documents of both the first technical document group and the second technical document group, as well as for calculating the sum, over all intermixed clusters, of a correction value proportional to the αth power (where $0<\alpha$) of the number of technical documents in each cluster, and dividing the sum by the calculated total number of clusters to calculate the similarity; and

a function, executed by the output means, for outputting the calculated similarity to recording means, to display means, or to communication means.

15. A similarity calculation program for calculating an index for judging technical similarity between technical document groups, which operates by means of information processing means for a similarity calculation device comprising technical document group input means for inputting the technical document groups, technical information input means for inputting technical information such as keywords, cluster analysis means for performing cluster analysis of the technical document groups by the technical information, similarity calculation means for calculating the total number of clusters and the number of intermixed clusters and

139

calculating the similarity, output means for outputting the calculated similarity, and information processing means capable of controlling the technical document group input means, the technical information input means, the cluster analysis means, the similarity calculation means, and the output means,

characterized in causing the information processing means to achieve:

a function, executed by the technical document group input means, for input of a first technical document group and a second technical document group for comparison;

a function, executed by the technical information input means, for input of the technical information such as keywords or IPC symbols;

a function, executed by the cluster analysis means, for retrieving technical documents containing the input technical information from technical documents contained in the first technical document group and the second technical document group, and for clustering the retrieved technical documents by each technical information;

a function, executed by the similarity calculation means, for calculating the total number of clusters obtained as a result of the cluster analysis and the number of intermixed clusters containing technical documents of both the first technical document group and the second technical document group, as well as for calculating the sum, over all intermixed clusters, of a correction value obtained by dividing the αth

140

power (where 0<α) of the number of technical documents in each cluster by a standardizing factor, and dividing the sum by the calculated total number of clusters to calculate the similarity; and

a function, executed by the output means, for outputting the calculated similarity to recording means, to display means, or to communication means.

16. The similarity calculation program according to Claim 15, further causing the information processing means to achieve a function, executed by the similarity calculation means, for using, as the standardizing factor, the average value of the number of technical documents in all clusters.

17. A similarity calculation program for calculating an index for judging technical similarity between technical document groups, which operates by means of information processing means for a similarity calculation device comprising technical document group input means for inputting the technical document groups, technical information input means for inputting technical information such as keywords, cluster analysis means for performing cluster analysis of the technical document groups by the technical information, similarity calculation means for calculating the total number of clusters and the number of intermixed clusters and calculating the similarity, output means for outputting the calculated similarity, and information processing means

capable of controlling the technical document group input

means, the technical information input means, the cluster

analysis means, the similarity calculation means, and the

output means,

characterized in causing the information processing means

to achieve:

a function, executed by the technical document group

input means, for input of a first technical document group and

a second technical document group for comparison;

a function, executed by the technical information input

means, for input of the technical information such as keywords

or IPC symbols;

a function, executed by the cluster analysis means, for

retrieving technical documents containing the input technical

information from technical documents contained in the first

technical document group and the second technical document

group, and for clustering the retrieved technical documents by

each technical information;

a function, executed by the similarity calculation means,

for calculating the total number of clusters obtained as a

result of the cluster analysis and the number of intermixed

clusters containing technical documents of both the first

technical document group and the second technical document

group, as well as for calculating the sum, over all intermixed

clusters, of a correction value proportional to the γth power

(where $0<\gamma$) of the probability of retrieving the m technical

documents from the first technical document group and the n

technical documents from the second technical document group, in order to perform correction according to the probability of the number of technical documents of the first technical document group and the second technical document group contained in each intermixed cluster obtained as a result of the cluster analysis, and dividing the sum by the calculated total number of clusters to calculate the similarity; and

a function, executed by the output means, for outputting the calculated similarity to recording means, to display means, or to communication means.

18. A similarity calculation program for calculating an index for judging technical similarity between technical document groups, which operates by means of information processing means for a similarity calculation device comprising technical document group input means for inputting the technical document groups, technical information input means for inputting technical information such as keywords, cluster analysis means for performing cluster analysis of the technical document groups by the technical information, similarity calculation means for calculating the total number of clusters and the number of intermixed clusters and calculating the similarity, output means for outputting the calculated similarity, and information processing means capable of controlling the technical document group input means, the technical information input means, the cluster analysis means, the similarity calculation means, and the

output means,

characterized in causing the information processing means to achieve:

a function, executed by the technical document group input means, for input of a first technical document group and a second technical document group for comparison;

a function, executed by the technical information input means, for input of the technical information such as keywords or IPC symbols;

a function, executed by the cluster analysis means, for retrieving technical documents containing the input technical information from technical documents contained in the first technical document group and the second technical document group, and for clustering the retrieved technical documents by each technical information;

a function, executed by the similarity calculation means, for calculating the total number of clusters obtained as a result of the cluster analysis and the number of intermixed clusters containing technical documents of both the first technical document group and the second technical document group, as well as for calculating the sum, over all intermixed clusters, of a correction value obtained by dividing, by a standardizing factor, the γth power (where $0<\gamma$) of the probability of retrieving the m technical documents from the first technical document group and the n technical documents from the second technical document group, in order to perform correction according to the probability of the number of

144

technical documents of the first technical document group and the second technical document group contained in each intermixed cluster obtained as a result of the cluster analysis, and dividing the sum by the calculated total number of clusters to calculate the similarity; and

a function, executed by the output means, for outputting the calculated similarity to recording means, to display means, or to communication means.

19. The similarity calculation program according to Claim 18, further causing the information processing means to achieve a function, executed by the similarity calculation means, for using, as the standardizing factor, the γth power (where $0<\gamma$) of the maximum value of the probability of retrieving the m technical documents from the first technical document group and the n technical documents from the second technical document group.

20. A similarity calculation program for calculating an index for judging technical similarity between technical document groups, which operates by means of information processing means for a similarity calculation device comprising technical document group input means for inputting the technical document groups, technical information input means for inputting technical information such as keywords, cluster analysis means for performing cluster analysis of the technical document groups by the technical information,

similarity calculation means for calculating the total number of clusters and the number of intermixed clusters and calculating the similarity, output means for outputting the calculated similarity, and information processing means capable of controlling the technical document group input means, the technical information input means, the cluster analysis means, the similarity calculation means, and the output means,

characterized in causing the information processing means to achieve:

a function, executed by the technical document group input means, for input of a first technical document group and a second technical document group for comparison;

a function, executed by the technical information input means, for input of the technical information such as keywords or IPC symbols;

a function, executed by the cluster analysis means, for retrieving technical documents containing the input technical information from technical documents contained in the first technical document group and the second technical document group, and for clustering the retrieved technical documents by each technical information;

a function, executed by the similarity calculation means, for calculating the total number of clusters obtained as a result of the cluster analysis and the number of intermixed clusters containing technical documents of both the first technical document group and the second technical document

group, as well as for calculating the sum, over all intermixed clusters, of a correction value proportional to the ζth power (where $0<\zeta$) of the ratio of a composition ratio N/M and an intermixing ratio n/m, for the composition ratio N/M of the number of technical documents N contained in the second technical document group to the number of technical documents M contained in the first technical document group and for the intermixing ratio n/m of the number of technical documents n of the second technical document group to the number of technical documents m of the first technical document group contained in each intermixed cluster obtained as a result of the cluster analysis, and dividing the sum by the calculated total number of clusters to calculate the similarity; and

a function, executed by the output means, for outputting the calculated similarity to recording means, to display means, or to communication means.

21. A similarity calculation program for calculating an index for judging technical similarity between technical document groups, which operates by means of information processing means for a similarity calculation device comprising technical document group input means for inputting the technical document groups, technical information input means for inputting technical information such as keywords, cluster analysis means for performing cluster analysis of the technical document groups by the technical information, similarity calculation means for calculating the total number

147

of clusters and the number of intermixed clusters and calculating the similarity, output means for outputting the calculated similarity, and information processing means capable of controlling the technical document group input means, the technical information input means, the cluster analysis means, the similarity calculation means, and the output means,

characterized in causing the information processing means to achieve:

a function, executed by the technical document group input means, for input of a first technical document group and a second technical document group for comparison;

a function, executed by the technical information input means, for input of the technical information such as keywords or IPC symbols;

a function, executed by the cluster analysis means, for retrieving technical documents containing the input technical information from technical documents contained in the first technical document group and the second technical document group, and for clustering the retrieved technical documents by each technical information;

a function, executed by the similarity calculation means, for calculating the total number of clusters obtained as a result of the cluster analysis and the number of intermixed clusters containing technical documents of both the first technical document group and the second technical document group, and calculating an expectation value for retrieving a

148

technical document of the first technical document group by

multiplying the probability of retrieving a technical document

of the first technical document group from among a technical

document group covering the first technical document group and

the second technical document group by the number of technical

documents contained in each intermixed cluster, and

calculating as an expectation value difference the difference

between the expectation value and the number of technical

documents of the first technical document group contained in

each intermixed cluster, as well as for calculating the sum,

over all intermixed clusters, of a correction value obtained

by setting the expectation value difference as negative

exponent for an arbitrary constant ξ (where $1<\xi$), and dividing

the sum by the calculated total number of clusters to

calculate the similarity; and

a function, executed by the output means, for outputting

the calculated similarity to recording means, to display means,

or to communication means.

22. A similarity calculation program for calculating an

index for judging technical similarity between technical

document groups, which operates by means of information

processing means for a similarity calculation device

comprising technical document group input means for inputting

the technical document groups, technical information input

means for inputting technical information such as keywords,

cluster analysis means for performing cluster analysis of the

technical document groups by the technical information, similarity calculation means for calculating the total number of clusters and the number of intermixed clusters and calculating the similarity, output means for outputting the calculated similarity, and information processing means capable of controlling the technical document group input means, the technical information input means, the cluster analysis means, the similarity calculation means, and the output means,

characterized in causing the information processing means to achieve:

a function, executed by the technical document group input means, for input of a first technical document group and a second technical document group for comparison;

a function, executed by the technical information input means, for input of the technical information such as keywords or IPC symbols;

a function, executed by the cluster analysis means, for retrieving technical documents containing the input technical information from technical documents contained in the first technical document group and the second technical document group, and for clustering the retrieved technical documents by each technical information;

a function, executed by the similarity calculation means, for calculating the total number of clusters obtained as a result of the cluster analysis and the number of intermixed clusters containing technical documents of both the first

150

technical document group and the second technical document group, and calculating the expectation value for retrieving a technical document of the first technical document group by multiplying the probability of retrieving a technical document of the first technical document group from among a technical document group covering the first technical document group and the second technical document group by the number of technical documents contained in each intermixed cluster, and calculating as an expectation value difference the difference between the expectation value and the number of technical documents of the first technical document group contained in each intermixed cluster, as well as for calculating the sum, over all intermixed clusters, of a correction value obtained by dividing the expectation value difference by the number of technical documents in each intermixed cluster and setting the divided expectation value difference as negative exponent for an arbitrary constant ξ (where $1<\xi$), and then dividing the sum by the calculated total number of clusters to calculate the similarity; and

a function, executed by the output means, for outputting the calculated similarity to recording means, to display means, or to communication means.

23. A similarity calculation method for calculating an index for judging technical similarity between technical document groups, using a similarity calculation device comprising technical document group input means for inputting

the technical document groups, technical information input means for inputting technical information such as keywords, cluster analysis means for performing cluster analysis of the technical document groups by the technical information, similarity calculation means for calculating the total number of clusters and the number of intermixed clusters and calculating the similarity, and output means for outputting the calculated similarity, comprising:

a process, executed by the technical document group input means, of inputting a first technical document group and a second technical document group for comparison;

a process, executed by the technical information input means, of inputting the technical information such as keywords or IPC symbols;

a process, executed by the cluster analysis means, of retrieving technical documents containing the input technical information from technical documents contained in the first technical document group and the second technical document group, and of clustering the retrieved technical documents by each technical information;

a process, executed by the similarity calculation means, of calculating the total number of clusters obtained as a result of the cluster analysis and the number of intermixed clusters containing technical documents of both the first technical document group and the second technical document group, and of calculating, as the similarity, the ratio of the number of intermixed clusters, containing technical documents

of both the first technical document group and the second technical document group, to the total number of clusters obtained as a result of the cluster analysis; and

a process, executed by the output means, of outputting the calculated similarity to recording means, to display means, or to communication means.

24. A similarity calculation method for calculating an index for judging technical similarity between technical document groups, using a similarity calculation device comprising technical document group input means for inputting the technical document groups, technical information input means for inputting technical information such as keywords, cluster analysis means for performing cluster analysis of the technical document groups by the technical information, similarity calculation means for calculating the total number of clusters and the number of intermixed clusters and calculating the similarity, and output means for outputting the calculated similarity, comprising:

a process, executed by the technical document group input means, of inputting a first technical document group and a second technical document group for comparison;

a process, executed by the technical information input means, of inputting the technical information such as keywords or IPC symbols;

a process, executed by the cluster analysis means, of retrieving technical documents containing the input technical

information from technical documents contained in the first

technical document group and the second technical document

group, and clustering the retrieved technical documents by

each technical information;

a process, executed by the similarity calculation means,

of calculating the total number of clusters obtained as a

result of the cluster analysis and the number of intermixed

clusters containing technical documents of both the first

technical document group and the second technical document

group, as well as calculating the sum, over all intermixed

clusters, of the product of a first correction value which

takes a value according to the number of technical documents

contained in each intermixed cluster and a second correction

value which takes a value according to the state of mixing of

technical documents of the first technical document group and

the technical documents of the second technical document group

in each intermixed cluster, and dividing the sum by the

calculated total number of clusters to calculate the

similarity; and

a process, executed by the output means, of outputting

the calculated similarity to recording means, to display means,

or to communication means.

25.   A similarity calculation method for calculating an

index for judging technical similarity between technical

document groups, using a similarity calculation device

comprising technical document group input means for inputting

the technical document groups, technical information input means for inputting technical information such as keywords, cluster analysis means for performing cluster analysis of the technical document groups by the technical information, similarity calculation means for calculating the total number of clusters and the number of intermixed clusters and calculating the similarity, and output means for outputting the calculated similarity, comprising:

a process, executed by the technical document group input means, of inputting a first technical document group and a second technical document group for comparison;

a process, executed by the technical information input means, of inputting the technical information such as keywords or IPC symbols;

a process, executed by the cluster analysis means, of retrieving technical documents containing the input technical information from technical documents contained in the first technical document group and the second technical document group, and clustering the retrieved technical documents by each technical information;

a process, executed by the similarity calculation means, of calculating the total number of clusters obtained as a result of the cluster analysis and the number of intermixed clusters containing technical documents of both the first technical document group and the second technical document group, as well as calculating the sum, over all intermixed clusters, of a correction value proportional to the αth power

155

(where 0<α) of the number of technical documents in each cluster, and dividing the sum by the calculated total number of clusters to calculate the similarity; and

a process, executed by the output means, of outputting the calculated similarity to recording means, to display means, or to communication means.

26. A similarity calculation method for calculating an index for judging technical similarity between technical document groups, using a similarity calculation device comprising technical document group input means for inputting the technical document groups, technical information input means for inputting technical information such as keywords, cluster analysis means for performing cluster analysis of the technical document groups by the technical information, similarity calculation means for calculating the total number of clusters and the number of intermixed clusters and calculating the similarity, and output means for outputting the calculated similarity, comprising:

a process, executed by the technical document group input means, of inputting a first technical document group and a second technical document group for comparison;

a process, executed by the technical information input means, of inputting the technical information such as keywords or IPC symbols;

a process, executed by the cluster analysis means, of retrieving technical documents containing the input technical

information from technical documents contained in the first technical document group and the second technical document group, and clustering the retrieved technical documents by each technical information;

a process, executed by the similarity calculation means, of calculating the total number of clusters obtained as a result of the cluster analysis and the number of intermixed clusters containing technical documents of both the first technical document group and the second technical document group, as well as calculating the sum, over all intermixed clusters, of a correction value obtained by dividing the αth power (where $0<\alpha$) of the number of technical documents in each cluster by a standardizing factor, and dividing the sum by the calculated total number of clusters to calculate the similarity; and

a process, executed by the output means, of outputting the calculated similarity to recording means, to display means, or to communication means.

27. The similarity calculation method according to Claim 26, wherein the similarity calculation means use, as the standardizing factor, the average value of the number of technical documents in all clusters.

28. A similarity calculation method for calculating an index for judging technical similarity between technical document groups, using a similarity calculation device

157

comprising technical document group input means for inputting the technical document groups, technical information input means for inputting technical information such as keywords, cluster analysis means for performing cluster analysis of the technical document groups by the technical information, similarity calculation means for calculating the total number of clusters and the number of intermixed clusters and calculating the similarity, and output means for outputting the calculated similarity, comprising:

a process, executed by the technical document group input means, of inputting a first technical document group and a second technical document group for comparison;

a process, executed by the technical information input means, of inputting the technical information such as keywords or IPC symbols;

a process, executed by the cluster analysis means, of retrieving technical documents containing the input technical information from technical documents contained in the first technical document group and the second technical document group, and clustering the retrieved technical documents by each technical information;

a process, executed by the similarity calculation means, of calculating the total number of clusters obtained as a result of the cluster analysis and the number of intermixed clusters containing technical documents of both the first technical document group and the second technical document group, as well as calculating the sum, over all intermixed

158

clusters, of a correction value proportional to the γth power (where $0<\gamma$) of the probability of retrieving the m technical documents from the first technical document group and the n technical documents from the second technical document group, in order to perform correction according to the probability of the number of technical documents of the first technical document group and the second technical document group contained in each intermixed cluster obtained as a result of the cluster analysis, and dividing the sum by the calculated total number of clusters to calculate the similarity; and

a process, executed by the output means, of outputting the calculated similarity to recording means, to display means, or to communication means.

29. A similarity calculation method for calculating an index for judging technical similarity between technical document groups, using a similarity calculation device comprising technical document group input means for inputting the technical document groups, technical information input means for inputting technical information such as keywords, cluster analysis means for performing cluster analysis of the technical document groups by the technical information, similarity calculation means for calculating the total number of clusters and the number of intermixed clusters and calculating the similarity, and output means for outputting the calculated similarity, comprising:

a process, executed by the technical document group input

159

means, of inputting a first technical document group and a second technical document group for comparison;

a process, executed by the technical information input means, of inputting the technical information such as keywords or IPC symbols;

a process, executed by the cluster analysis means, of retrieving technical documents containing the input technical information from technical documents contained in the first technical document group and the second technical document group, and clustering the retrieved technical documents by each technical information;

a process, executed by the similarity calculation means, of calculating the total number of clusters obtained as a result of the cluster analysis and the number of intermixed clusters containing technical documents of both the first technical document group and the second technical document group, as well as calculating the sum, over all intermixed clusters, of a correction value obtained by dividing, by a standardizing factor, the γth power (where $0<\gamma$) of the probability of retrieving the m technical documents from the first technical document group and the n technical documents from the second technical document group, in order to perform correction according to the probability of the number of technical documents of the first technical document group and the second technical document group contained in each intermixed cluster obtained as a result of the cluster analysis, and dividing the sum by the calculated total number

160

of clusters to calculate the similarity; and

a process, executed by the output means, of outputting the calculated similarity to recording means, to display means, or to communication means.

30. The similarity calculation method according to Claim 29, wherein the similarity calculation means use, as the standardizing factor, the γth power (where $0<\gamma$) of the maximum value of the probability of retrieving the m technical documents from the first technical document group and the n technical documents from the second technical document group.

31. A similarity calculation method for calculating an index for judging technical similarity between technical document groups, using a similarity calculation device comprising technical document group input means for inputting the technical document groups, technical information input means for inputting technical information such as keywords, cluster analysis means for performing cluster analysis of the technical document groups by the technical information, similarity calculation means for calculating the total number of clusters and the number of intermixed clusters and calculating the similarity, and output means for outputting the calculated similarity, comprising:

a process, executed by the technical document group input means, of inputting a first technical document group and a second technical document group for comparison;

a process, executed by the technical information input means, of inputting the technical information such as keywords or IPC symbols;

a process, executed by the cluster analysis means, of retrieving technical documents containing the input technical information from technical documents contained in the first technical document group and the second technical document group, and clustering the retrieved technical documents by each technical information;

a process, executed by the similarity calculation means, of calculating the total number of clusters obtained as a result of the cluster analysis and the number of intermixed clusters containing technical documents of both the first technical document group and the second technical document group, as well as calculating the sum, over all intermixed clusters, of a correction value proportional to the ζth power (where $0<\zeta$) of the ratio of a composition ratio N/M and an intermixing ratio n/m, for the composition ratio N/M of the number of technical documents N contained in the second technical document group to the number of technical documents M contained in the first technical document group and for the intermixing ratio n/m of the number of technical documents n of the second technical document group to the number of technical documents m of the first technical document group contained in each intermixed cluster obtained as a result of the cluster analysis, and dividing the sum by the calculated total number of clusters to calculate the similarity; and

162

a process, executed by the output means, of outputting the calculated similarity to recording means, to display means, or to communication means.

32. A similarity calculation method for calculating an index for judging technical similarity between technical document groups, using a similarity calculation device comprising technical document group input means for inputting the technical document groups, technical information input means for inputting technical information such as keywords, cluster analysis means for performing cluster analysis of the technical document groups by the technical information, similarity calculation means for calculating the total number of clusters and the number of intermixed clusters and calculating the similarity, and output means for outputting the calculated similarity, comprising:

a process, executed by the technical document group input means, of inputting a first technical document group and a second technical document group for comparison;

a process, executed by the technical information input means, of inputting the technical information such as keywords or IPC symbols;

a process, executed by the cluster analysis means, of retrieving technical documents containing the input technical information from technical documents contained in the first technical document group and the second technical document group, and clustering the retrieved technical documents by

163

each technical information;

a process, executed by the similarity calculation means, of calculating the total number of clusters obtained as a result of the cluster analysis and the number of intermixed clusters containing technical documents of both the first technical document group and the second technical document group, and calculating an expectation value for retrieving a technical document of the first technical document group by multiplying the probability of retrieving a technical document of the first technical document group from among a technical document group covering the first technical document group and the second technical document group by the number of technical documents contained in each intermixed cluster, and calculating as an expectation value difference the difference between the expectation value and the number of technical documents of the first technical document group contained in each intermixed cluster, as well as calculating the sum, over all intermixed clusters, of a correction value obtained by setting the expectation value difference as negative exponent for an arbitrary constant ξ (where $1<\xi$), and dividing the sum by the calculated total number of clusters to calculate the similarity; and

a process, executed by the output means, of outputting the calculated similarity to recording means, to display means, or to communication means.

33. A similarity calculation method for calculating an

index for judging technical similarity between technical document groups, using a similarity calculation device comprising technical document group input means for inputting the technical document groups, technical information input means for inputting technical information such as keywords, cluster analysis means for performing cluster analysis of the technical document groups by the technical information, similarity calculation means for calculating the total number of clusters and the number of intermixed clusters and calculating the similarity, and output means for outputting the calculated similarity, comprising:

a process, executed by the technical document group input means, of inputting a first technical document group and a second technical document group for comparison;

a process, executed by the technical information input means, of inputting the technical information such as keywords or IPC symbols;

a process, executed by the cluster analysis means, of retrieving technical documents containing the input technical information from technical documents contained in the first technical document group and the second technical document group, and clustering the retrieved technical documents by each technical information;

a process, executed by the similarity calculation means, of calculating the total number of clusters obtained as a result of the cluster analysis and the number of intermixed clusters containing technical documents of both the first

technical document group and the second technical document group, and calculating the expectation value for retrieving a technical document of the first technical document group by multiplying the probability of retrieving a technical document of the first technical document group from among a technical document group covering the first technical document group and the second technical document group by the number of technical documents contained in each intermixed cluster, and calculating as an expectation value difference the difference between the expectation value and the number of technical documents of the first technical document group contained in each intermixed cluster, as well as calculating the sum, over all intermixed clusters, of a correction value obtained by dividing the expectation value difference by the number of technical documents in each intermixed cluster and setting the divided expectation value difference as negative exponent for an arbitrary constant ξ (where $1<\xi$), and then dividing the sum by the calculated total number of clusters to calculate the similarity; and

a process, executed by the output means, of outputting the calculated similarity to recording means, to display means, or to communication means.